

Exploratory Data Analysis and Predictive Modeling for Diabetes Detection

Tommi Bimbato

March 24, 2025

Abstract

This project focuses on exploratory data analysis (EDA) and predictive modeling for identifying potential diabetes in patients.

The study is conducted as part of the course "Programming and Database - Data Science" at the University of Verona (UNIVR) for the academic year 2024/2025.

The dataset used for this project is `diabetes_unclean.csv`, which contains raw health data from various patients. The project involves data cleaning, exploratory data analysis, and predictive modeling to estimate the likelihood of diabetes in patients. The sources of the information can be found in the 'readme.md' file in the repository.

The repository is structured as follows:

- `data/`: Contains the raw dataset ('`diabetes_unclean.csv`') and any processed datasets.
- `notebooks/`: Jupyter notebooks for data cleaning, EDA, and modeling.
- `src/`: Python scripts for data processing and model development.
- `results/`: Outputs, visualizations, and model evaluation results.

This project is purely educational and part of the coursework for the **Programming and Database** course. It should not be considered for scientific or professional use.